DISCRIMINATION ABILITY OF NEURAL NETWORK AND MAXIMUM LIKELIHOOD CLASSIFIERS

Xue-Hua LIU*, **Andrew SKIDMORE****, **Henk van OOSTEN*****
International Institute for Aerospace Survey and Earth Science (ITC)
P. O. Box 6, 7500 AA Enschede, The Netherlands
*E-mail: liuxh@itc.nl
**E-mail: skidmore@itc.nl
***E-mail: oost@itc.nl

KEY WORDS: Remote Sensing, Algorithms, Neural Network, Maximum Likelihood, Classification, Accuracy.

ABSTRACT

Artificial neural networks have been taken as powerful tools for pattern recognition and data analysis. However, the previous classification results show differences between the neural network classifier (NNC) and the statistical methods, such as maximum likelihood classifier (MLC). The reasons for the differences in classification accuracy were not completely understood. Much research has explored the behaviour of NNCs, including inputs, number of hidden layer(s), number of nodes, size of sample sets, training parameters. However, little work has been done on the effect of the overlap degree of patterns in feature space on a NNC. This study analyses the effect of the overlap degree of classes in feature space on performance of a standard backpropagation NNC and compares the results with MLC. Two data sets (i.e., the simulated data sets with different overlap degrees and the remotely sensed imagery with more complicate overlap pattern) are used to test the performance of a neural network. One surprising result is that NNC failed to discriminate two non-overlapping classes. Another interesting result is that NNC classifies class2 better while MLC classifies class1 better. "Z" statistics is carried out to compare two classifiers and results show that NNC is significantly better than MLC in the case of using simulated data at 95% C.I., but not significantly different with MLC in the case of using actual image in the study.

1 INTRODUCTION

Artificial neural networks have successfully classified remotely sensed data (Hepner et al., 1990, Zhuang et al., 1994, Weeks and Gaston, 1997). There are significant differences between neural network and statistical classifiers, such as the maximum likelihood classifier (Bischof et al., 1992, Chen et al., 1993, Solaiman and Mouchot, 1994, Paola and Schowengerdt, 1995, Weeks and Gaston, 1997): (1) neural networks make no assumptions about the input data; (2) neural networks form non-linear decision boundaries in the feature space; (3) neural networks are robust when presented with partially incomplete or incorrect input patterns; (4) neural networks can generalise inputs.

In a comprehensive review, Paola and Schowengerdt (1995) concluded that neural network classifiers yield similar (or slightly higher) accuracy when compared to conventional MLC (Hepner et al., 1990, Key et al., 1990, Bischof et al., 1992, Kanellopollos et al., 1992, Paola and Schowengerdt, 1994). As a result of the marginal improvement in mapping accuracy by neural network classifiers, Skidmore et al. (1997) recommended maximum likelihood classifiers, as they are easier to use. Indeed, some authors have found maximum likelihood classifiers give a higher mapping accuracy than neural networks (Benediktsson et al., 1990, Solaiman and Mouchot, 1994).

Fierens et al. (1994) were unable to understand why classifiers have differences in accuracy. One of the reasons could be that the experimental set-ups are not comparable. For instance, texture measures have been used with NNCs, but not utilised by conventional classifiers (Hepner et al., 1990, Bischof et al., 1992, Paola and Schowengerdt, 1994, Skidmore et al., 1997). Another reason for differences in classification accuracy is that the assumptions of a classifier may be better met by a particular image data set. For example, Key et al. (1990) theorised that NNC avoids assumptions of statistical normality, and has greater flexibility to classify indistinct classes. However, Benediktsson et al. (1990) used data from a random number generator with normalised distribution and found the accuracy of the maximum likelihood classifier was higher than a neural network. Thus it may be assumed that the normalised distribution allows the maximum likelihood method to perform well.

The performance of a backpropagation neural network is affected by many factors. A number of researchers have focused on exploring the behaviour of neural networks by adjusting factors such as the input data, number of hidden layers, number of nodes, as well as different training parameters such as momentum, learning rate and number of epochs (Benediktsson et al., 1990, Heermann and Khazenie, 1990, Zhuang et al., 1994, Ardo et al., 1997, Skidmore et al., 1997). However, no authors have investigated the effect of the degree of overlap between classes in feature space on the performance of neural network and conventional classifiers. The aims of this paper are, by using different degrees of overlap between classes, to compare: (1) the accuracy of the NNC in response to different degrees of overlap in simulated data sets; (2) the performance of NNC and MLC under different levels of overlap in feature space; (3) the performance of the different classifiers on a real image data set.

2 BACKGROUND AND ASSUMPTIONS OF MLC AND NNC

The maximum likelihood classifier is the most commonly used supervised classification method. The decision rule is defined by the multidimensional normal distribution around a class mean (see figure 1a). Consequently, multi-modal distributed data will cause an incorrect classification. In addition, overlapping decision boundaries in feature space will be problematic, especially if the training data do not physically overlap, but the decision boundaries do overlap (e.g. Skidmore et al., 1988, Fierens et al., 1994).

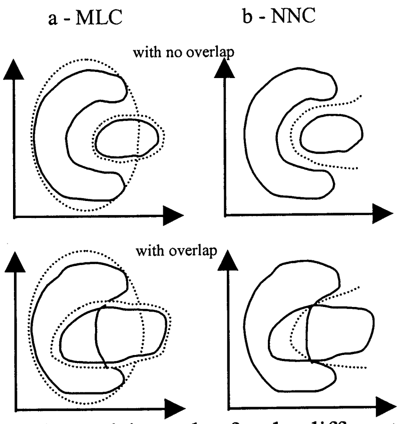

Figure 1. Decision rules for the different classifiers in a two-dimensional feature space

Neural network classifiers recognise spectral patterns by learning training sets. They contain three or more layers of nodes viz. an input layer, (a) hidden layer(s) and an output layer. The error between the network output and the target is reduced by adjusting all weights of the network until the system error falls below a user specified threshold. After training, the neural network system fixes all weights and maintains the original learning parameters. The classification process calculates the output of each pixel using the parameters learnt from the training phase, and then decides the class of the pixel. Solaiman and Mouchot (1994) mentioned that the multi-layer perceptron is a decision-surface based classifier. According to Skidmore et al. (1988), it is possible to explain the separation of classes by MLC in feature space (see figure 1a at the left side). Richards (1995), therefore, hypothesised that a hyperplane decision surface between two different classes may also be created for NNC that can divide the pattern space into different regions (see figure 1b at the right side).

3 METHODS

3.1 Data sets

Two classes were simulated varying from a condition of no overlap, to a condition of overlap, in order to test the effect of feature separability and overlap degree on NNC and MLC. The classes were generated for a random normal distribution in three bands. The total number of pixels per class (i.e., 5000 pixels) was first defined. All these pre-conditions (such as two classes, three bands and fixed pixels per class) are used to easily compare the classifiers. Figure 2 shows the feature space of the simulated data sets with two classes in three dimensions.

In addition to the synthetic data sets, the performance of two classifiers were tested using remotely sensed imagery (viz. Landsat TM and SPOT-pan acquired in 1995 and 1997 respectively) over the Lemeleberg region of the Netherlands. The images were geometrically rectified and geo-referenced to a common pixel size of 10 by 10 meters. The sub-images contain five ground cover classes.

3.2 Defining separability and degree of overlap of the classes

Accuracy of classification depends on the separability of the classes in feature space. As two classes become further apart, they have less overlap and may be classified with a greater accuracy. Two measures are used in this study: the Jeffries-Matusita (JM) distance (ERDAS, 1991) and the simplified Skidmore et al. non-parametric test of overlap (Skidmore et al., 1988).

Mathematical separability is normally used to discard classes with little contribution to a classification (Richards, 1995). The JM Distance is a parametric measure of the average distance between the density function of two classes. For normally distributed classes, JM Distance may be defined as:

$$J_{ij} = 2(1 - e^{-B}) \tag{1}$$

where B is the Bhattacharyya Distance:

$$B = 1/8(\mu_i - \mu_j)^t \{\frac{C_i + C_j}{2}\}^{-1}(\mu_i - \mu_j) + 1/2\ln\{\frac{|(C_i + C_j)2|}{\sqrt{|C_i|}\sqrt{|C_j|}}\} \tag{2}$$

i and j are two classes being compared, μ_i and μ_j are the mean vectors of the two classes and C_i and C_j are the variance-covariance matrices of the two classes. The JM distance ranges from 0 where the two classes completely overlap to 2 where the two classes completely are separate from each other.

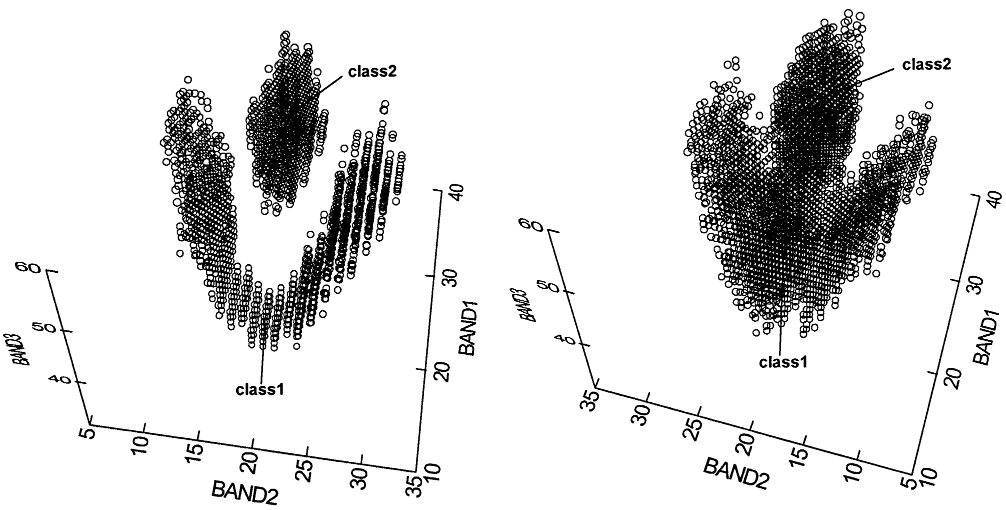

Figure 2. The feature space of the simulated data sets in three dimensions

Skidmore et al. (1988) developed a general algorithm to quantify the degree of overlap of classes. It is a non-parametric test of overlap that does not depend on statistical parameters. The Skidmore et al. $R_i(f)$ value ranges from 0 to 1, where 0 equates to complete overlap, while 1 means the two classes have no overlap at all, that is completely separate from each other. In this study, a simplified Skidmore et al. non-parametric was used:

$$R_i = F_i(X) / N_i \tag{3}$$

where $F_i(X)$ is defined as the frequency of pixels in the training set belonging to class i, N_i is the total number of pixels of the training set of class i, and R_i is the proportion of $F_i(X)$ and N_i, which is used to indicate the overlap degree of class i with other classes. If R_i=1, this class has no overlap with other classes, while if R_i<1, the class has overlap with other classes.

3.3 Configuring the neural network

The neural networks with three layers (1 input layer, 1 hidden layer and 1 output layer) were constructed with varying number of nodes. A 3-5-2 network, e.g. three input nodes, five hidden nodes and two output nodes was applied to the simulated data sets, while a 3-5-5 network was developed for the real image data. In order to optimise training, a small learning rate (0.001) and momentum (0.01) were used.

3.4 Testing the neural network classifier using simulated data sets

A number of different conditions are tested using the neural network classifier, including two overlap situations, the size of the training sample sets (i.e., 200, 400, 800, 1600 and 2500), and the system error. The test sets per class are the pixels (i.e., 4800, 4600, 4200, 3400 and 2500) remained after excluding the training samples from 5000 pixels per class. The overall accuracy was calculated and used for evaluating the influence of different experiment conditions on NNC.

3.5 Comparing the classifiers using both simulated and remotely sensed imagery and 'Z' statistics

Both simulated data and remotely sensed imagery were classified using the NNC and MLC in order to test the effect of feature separability and overlap degree on classification accuracy. The same training samples and testing sets were used for two classifiers.

The measure of the agreement between two separate error matrices, called KHAT, can be used to test the significant difference between two classified images (Congalton et al., 1983, Foody, 1992). A pairwise test of significance can be performed between two independent KHATs using the normal curve deviate to determine if the two error matrices are significantly different (Cohen, 1960). The classifiers tested in this study are compared using KHAT measure, or called 'z' statistic (see equation (4)). K_1 and K_2 represents the KHAT values calculated from two different classifications. $\delta_{K_1}^{\ 2}$ and $\delta_{K_2}^{\ 2}$ are the variances from K_1 and K_2 respectively. For pairwise comparison, a null hypothesis can be defined to test whether the K values for two classifications differ. The null hypothesis is rejected using the normal curve deviate statistic (z) for α=0.05 if Z_t>1.96. Note that the compared result from NNC is the one at the system error of 0.01 for 'no overlap' situation and 0.035 for 'overlap' situation.

$$Z_t = \frac{K_1 - K_2}{\sqrt{\delta_{K_1}^{\ 2} + \delta_{K_2}^{\ 2}}} \tag{4}$$

4 RESULTS

4.1 Measuring the separability and overlap degree for the simulated data set

The separability (JM distance) and degree of overlap (R_i value) are detailed in table 1. The JM distances are approximately 1.52 under the no overlap situation and approximately 1.51 under the overlap situation. Similarity of these results indicates that the two classes can not be completely separated under both situations. Nevertheless, the simplified-R_i values are all equal to 1 under the no overlap situation which exactly show two classes have no overlap, and vary from 0.969 to 0.995 under the overlap situation.

Sample size per class	No overlap		Overlap	
	JM distance	Simplified Ri	JM distance	Simplified Ri
200	1.519	Rc1=1.000 Rc2=1.000	1.509	Rc1=0.995 Rc2=0.995
400	1.518	Rc1=1.000 Rc2=1.000	1.509	Rc1=0.990 Rc2=0.988
800	1.517	Rc1=1.000 Rc2=1.000	1.509	Rc1=0.974 Rc2=0.970
1600	1.518	Rc1=1.000 Rc2=1.000	1.509	Rc1=0.975 Rc2=0.972
2500	1.518	Rc1=1.000 Rc2=1.000	1.509	Rc1=0.971 Rc2=0.969

Table 1. JM distance and simplified R_i value under different sampling schemes

4.2 Performance of the different classifiers with simulated data set

For the 'no overlap' condition, Figure 3 summarises the change in classification accuracy in response to a varying system error and training set size. Similarly, Figure 4 summarises the results under the 'overlap' condition. Note that 0.1, 0.075, 0.05, 0.035, 0.025, 0.01 are the system error levels applied to NNC, while MLC refers to the results from maximum likelihood classifier. Under overlap situation, training NNC became difficulty and stopped at the system error of 0.035. Figures 3a and 4a show that the number of correctly classified pixels in class1 increases as the system error reduces and the size of the training sample set enlarges. However, the number of correctly classified pixels in class2 varies under the similar conditions (figure 3b, 4b). The number of pixels correctly classified by NNC for both two classes under the no overlap situation is higher than under the overlap situation. Figures 3c and 4c show that the overall classification accuracy is higher under the small system error and non-overlap situation. The surprising result is that the neural network fails to classify two 'no overlap' classes with an accuracy of 100 per cent.

The results from NNC are compared with the traditional MLC and the classified images from MLC and NNC with the system error 0.01 for no overlap situation and 0.035 for overlap situation are shown in figure 5. Under both the non-overlap and overlap situations, MLC and NNC have the different results on classifying two classes. NNC classifies class2 with a higher accuracy while MLC classifies class1 better. When viewed as classified images, figure 5 highlights the differences between MLC and NNC.

4.3 Performance of NNC and MLC using Landsat TM and SPOT-pan imagery

The performance of the NNC and MLC was evaluated using the combined Landsat TM and SPOT-pan images comprising three channels: TM band 2 and 4, and SPOT-pan. Five ground covers were recognised in the field, viz. Forest (F), pasture (P), heathland (H), arableland (A) and buildup area (B).

Table 2 details the separability (JM distance) of pairs of classes as well as the overlap degree (R_i value) of each class with other classes based on the sample sets. The highest separability is between pasture and forest, followed by forest and arableland, pasture and heathland, pasture and buildup. The R_i values show forest and pasture do not overlap with the other classes in the training sample sets, while the other three classes have some overlap with each other. The buildup-area class has the highest degree of overlap with other classes followed by the arableland class. Table 3 details the accuracy matrices for the classification of the real image by the NNC and MLC. NNC has the highest overall accuracy, being a little higher than MLC. The pasture class was classified with the highest producer accuracy and the buildup-area class with the lowest producer accuracy. NNC did not separate the forest and pasture classes from the other three classes with 100 per cent accuracy although the R_i values show that these two classes have no overlap.

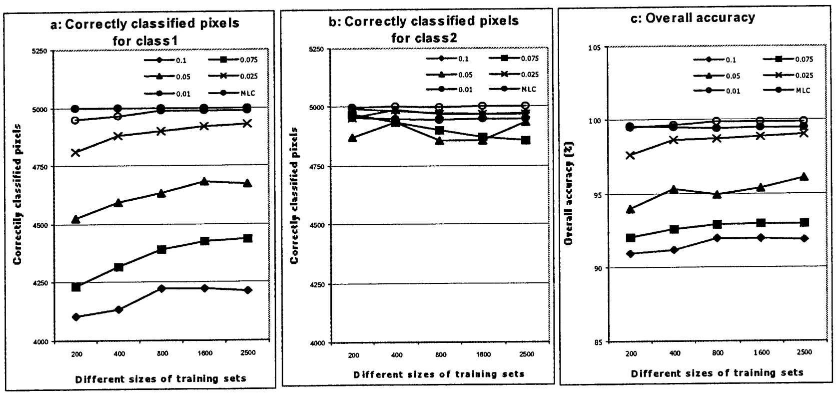

Figure 3. Results of NNC and MLC under no overlap situation

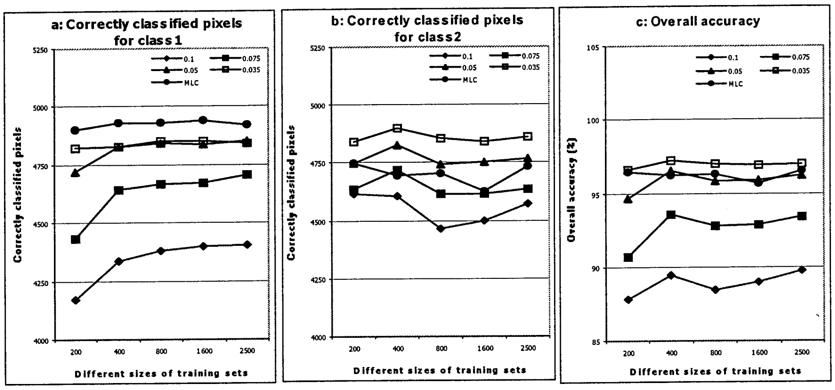

Figure 4. Results of NNC and MLC under overlap situation

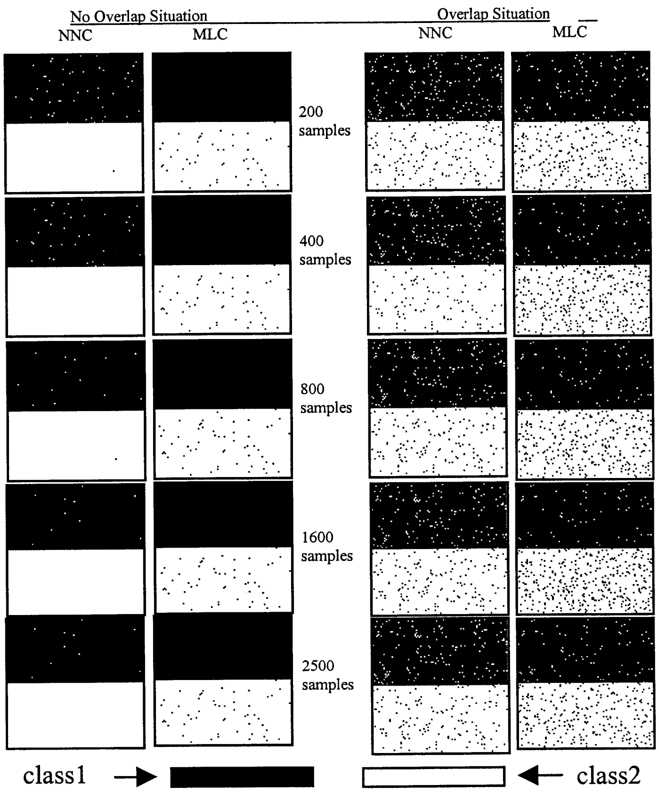

Figure 5. The classified images from NNC and MLC for two overlap situations

4.4 'Z' statistic test to compare NNC and MLC

Table 4 details whether there is a statistically significant difference between the accuracy of the images produced by NNC and MLC classifiers, shown by all 'z' values. Under the no overlap situation with the simulated data sets, NNC and MLC show no significant difference with small samples but significant difference with large samples. With the overlapping classes, NNC has a significantly higher accuracy than MLC mostly. In the case of the Lemeleberg image classification, there is no significant difference between NNC and MLC although NNC has a slightly higher overall accuracy than MLC.

	JM					Ri	Training Samples
	F	P	H	A	B		
F	0	1.609	1.514	1.595	1.536	R_F=1	285
P		0	1.569	1.538	1.551	R_P=1	473
H			0	1.545	1.506	R_H=0.937	300
A				0	1.519	R_A=0.913	480
B					0	R_B=0.897	312

Table 2. The JM distances and Simplified R_i value

5 DISCUSSION

According to Richards (1995), a JM distance of 2.0 implies that the classes may be classified with an accuracy of 100 per cent. Therefore, the JM values in table 1 show that there is overlap for both the 'no overlap' and the 'overlap' situations. However, if we look at the simplified Skidmore's R_i values, the 'no overlap' situation has a R_i value equal to 1.0, correctly showing there is no overlap between two classes. Thus, the JM informs a user on how well two classes may be classified, but gives no information about the degree of overlap.

Classes from test sets (N=5319)*	Classes from NNC classification					Producer accuracy	Overall accuracy (%)
	F	P	H	A	B		
F	1260	25	113	0	0	0.90	
P	0	1184	35	0	11	0.96	
H	30	65	645	0	3	0.87	
A	0	20	0	920	232	0.78	
B	0	153	112	137	374	0.48	82.40
Classes from test sets	Classes from MLC classification					Producer accuracy	Overall accuracy (%)
	F	P	H	A	B		
F	1180	0	218	0	0	0.84	
P	0	1190	39	0	1	0.97	
H	11	57	672	0	3	0.90	
A	0	8	0	848	316	0.72	
B	0	147	103	89	437	0.56	81.35

* N --- total sample number for testing.

Table 3. Overall accuracies of classification using two classifiers

One important result of this study is that NNC failed to classify the two 'no overlapping' classes in the feature space although NNC produces a higher classification accuracy than MLC. Since a decision hyper plane can be formed by the NNC between two classes (Richards 1995), theoretically, NNC should be able to classify two non-overlapping classes with overall accuracy of 100 per cent, if it can be well trained. However, training sets represent only part of the whole data set.

Sample size	Z for 'no overlap' situation	Z for 'overlap' situation
200	0.1008	0.5801
400	1.8062	4.0966*
800	4.9620*	2.7983*
1600	5.4303*	4.7970*
2500	5.4303*	1.7193
Remotely sensed imagery	1.2872	

* with significant difference at 95% C.I.

Table 4. Pairwise comparisons between MLC and NNC

Turning now to the classification accuracy of individual classifiers, it has been shown that MLC classifies class1 with a high accuracy but has more misclassified pixels in class2, while NNC obtains the highest accuracy with class2 and produces more misclassified pixels in class 1. MLC is a parametric method which utilises the mean and standard deviation of each band. Therefore, as class1 covers a large spectral range, MLC can classify class1 better using the shortest Mahalanobis distance and also decides its lower accuracy in classifying class2 due to its decision rule (see figure 1a). However, it is not completely clear why non-parametric NNC can not classify class1 as good as class2 (see figure 1b). One reason could be the size of training sample set because class1 has a wide range of brightness values. A similar result was also obtained by Downey et al. (1992), who found that NNC achieved accuracies of 90.59% and 12.49% for woodland and cropland classes respectively compared to 34.99% and 66.46% for the same two classes using MLC. It implies that integrating two methods together may get higher classification accuracy because two methods can compensate each other.

When remotely sensed imagery were investigated, the high simplified R_i values for the forest and pasture classes indicated that these two classes do not overlap with any other class in feature space. However, the classification accuracies for these two classes are lower than 100 per cent. The classifiers (NNC and MLC) are able to create decision surfaces that can discriminate the classes in feature space. As the NNC is a non-parametric classifier, it is surprising that the decision surface could not better discriminate the non-overlapping classes, similar as the results from the simulated data set. It may be due to the sample set can not represent the whole data set. The 'buildup area' class has a high level of overlap with other classes in feature space, e.g., has the lowest R_i-value of 0.897 and JM-value of 1.506, and also the lowest classification accuracy.

6 CONCLUSION

It is concluded from the study that overlap of classes in feature space produces misclassification by both NNC and MLC, for both simulated data and remotely sensed imagery. Experiments based on the simulated data sets shows that NNC and MLC have different accuracies in mapping the two classes. The well-trained neural network can classify the simulated data sets significantly better than MLC. Classification of remotely sensed imagery (Landsat TM and SPOT-pan) in Lemeleberg, the Netherlands, shows that there is no significant difference between NNC and MLC.

ACKNOWLEDGEMENT

We would like to thank Ms Hong Yang and Mr Lucas H. Broekema for their help in executing the neural network package in PCI software in HP-workstation. We also like to thank Mr K. Bronsveld for his comments on paper and providing the web-sites of the neural network software, thank Mr G. Reinink for his kind help.

REFERENCE

Ardo, J., Pilesjo, P. and Skidmore, A. K., et al. 1997, Neural network, multitemporal Landsat Thematic Mapper data and topographic data to classify forest damages in the Czech Republic. *Canadian Journal of Remote Sensing* , **23**(3), pp. 217-229.

Benediktsson, J. A., Swain, P.H., Ersoy, O. K., et al. and Hong, D., 1990, Classification of very high dimensional data using neural networks. *Proceedings of the International Geoscience and Remote Sensing Symposium IGARSS'90, the University of Maryland College park, MarylandD., USA, 20-24 May 1990* (New York: Institute of Electrical and Electronics Engineers), pp. 1269-1272.

Bischof et al., H., Schneider, W., and Pinz, A. J., 1992, Multispectral classification of Landsat-images using neural networks. *IEEE Transactions on Geosceience and Remote Sensing* , **30**(3), pp. 482-490.

Chen et al., K. S., Tzeng, Y. C., Chen, C. F., Kao, W. L. and Ni, C. L., 1993, Classification of multispectral imagery using dynamic learning neural network. *Proceedings of the International Geoscience and Remote Sensing Symposium IGARSS'93, Kogakuin University, Tokyo, Japan, 18-21 August 1993* .(Piscataway, NJ: Institute of Electrical and Electronics Engineers), pp. 896-898.

Cohen, J., 1960, A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* , XX(1), pp. 37-46.

Congalton, et al.R. G., Oderwald, R. G., and Mead, R. A., 1983, Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques. *Photogrammetric Engineering & Remote Sensing* , **49**(12), pp. 1671-1678.

Downey et al., I. D., Power, C. H., Kanellopoulos, I., and Wilkinson, G. G., 1992, A performance comparison of Landsat TM land cover classification based on neural network techniques and traditional maximum likelihood algorithms and minimum distance algorithms. *Proceedings of the 1992 18th Annual Conference of the UK Remote Sensing Society, , Dundee, Scotland, 15-17 Sepetember1992,* pp. 518-528.

ERDAS IMAGINE, 1991, ERDAS Imagine field guide: Part 1,2,3: ERDAS Imagine version 8.2 for Windows NT and Windows 95. 3rd edn. Atlanta.

Fierens et al, F., Kanellopoulos, I., Wilkinson, G. G., and Megier, J., 1994, Comparison and visualisation of feature space behaviour of statistical and neural classifiers of Satellite imagery. *Proceedings of the International Geoscience and Remote Sensing Symposium IGARSS'94, Pasadena, California, USA, 8-12 August 1994* (Piscataway, NJ: Institute of Electrical and Electronics Engineers), pp. 1880-1882..

Foody, G. M., 1992, On the compensation for chance agreement in image classification accuracy assessment. *Photogrammetric Engineering & Remote Sensing* , **58**(10), pp. 1459-1460.

Hepner et al., G. F., Logan, T., Ritter, N. and Bryant, N., 1990, Artificial neural network classification using a minimal training set: comparison to conventional supervised classification. *Photogrammetric Engineering and Remote Sensing*, **56**(4), pp. 469-473.

Heermann, P. D., and Khazenie, N., 1990, Application of neural networks for classification of multi-source multi-spectral remote sensing data. *Proceedings of the International Geoscience and Remote Sensing Symposium IGARSS'90, the University of Maryland College park, Maryland, USA, 20-24 May 1990, College park, MD.* (New York: Institute of Electrical and Electronics Engineers), pp. 1273-1276.

Kanellopoulos et al., I., Varfis, A., and Wilkinson, G. G., 1992, Land-cover discrimination in SPOT HRV imagery using an artificial neural network - a 20-class experiment. *International Journal of Remote Sensing* , **13**(5), pp. 917-924.

Key et al., J., Maslanik, J. A. and Schweiger, A. J., 1990, Neural network vs. maximum likelihood classifications of spectral and textural features in visible, thermal, and passive microwave data. *Proceedings of the International Geoscience and Remote Sensing Symposium IGARSS'90, the University of Maryland College park, Maryland, USA, 20-24 May 1990 College park, MD.*(New York: Institute of Electrical and Electronics Engineers), pp. 1277-1280.

Paola, J. D., and Schowengerdt, R. A., 1994, Comparisons of neural networks to standard techniques for image classification and correlation. *Proceedings of the International Geoscience and Remote Sensing Symposium IGARSS'94, Pasadena, California, USA, 8-12 August 1994* (Piscataway, NJ: Institute of Electrical and Electronics Engineers), pp. 1404-1406.

Paola, J. D., and Schowengerdt, R. A., 1995, A review and analysis of backpropagation neural networks for classification of remotely-sensed multi-spectral imagery. *International Journal of Remote Sensing*, **16**(16), pp. 3033-3058.

Richards, J. A., 1995, Remote sensing digital image analysis - An Introduction. 2nd edn (New York,: Springer-Verlag).

Skidmore, A. K., Forbes, G. W., and Carpenter, D. J., 1988, Non-parametric test of overlap in multispectral classification. *International Journal of Int. J. Remote Sensing,* **9**(4), pp. 777-785.

Skidmore et al., A. K., Turner, B. J., Brinkhof, W., and Knowles, E., 1997, Performance of a neural network: mapping forests using GIS and remotely sensed data. *Photogrammetric Engineering & Remote Sensing*, **63**(5), pp. 501-514.

Solaiman, B. and Mouchot, M. C., 1994, A comparative study of conventional and neural network classification of multispectral data. *Proceedings of the International Geoscience and Remote Sensing Symposium IGARSS'94, Pasadena, California, USA, 8-12 August 1994* (Piscataway, NJ: Institute of Electrical and Electronics Engineers), pp. 1413-1415..

Weeks, P. J. D., and Gaston, K. J., 1997, Image analysis, neural networks and the taxonomic impediment to biodiversity studies. *Biodiversity and Conservation*, **(6)**, pp. 263-274.

Zhuang et al., X., Engel, B. A., Lozano-Garcia, D. F., Fernandez, R. N., and Johannsen, C. J., 1994, Optimisation of training data required for neuro-classification. *International Journal of Remote Sensing*, **15**(16), pp. 3271-3277.